

ZERO-SHOT COLOR CONSTANCY BY ESTIMATING ALBEDO

Gabriele Canesi

Marco Buzzelli

Simone Bianco

Raimondo Schettini

University of Milan-Bicocca

ABSTRACT

Computational color constancy (CC) aims to recover the true colors of a scene despite changes in illumination. While recent cross-camera deep learning methods have made impressive progress, their training relies on relatively small RAW datasets. In this work, we introduce a zero-shot color constancy framework grounded in intrinsic image decomposition (IID). Our method estimates the albedo and infers the illumination map from a pre-trained IID model, followed by chromaticity clustering. This design allows us to avoid any fine-tuning on CC datasets. Experiments on the Gehler-Shi and NUS-8 datasets show that our approach achieves performance comparable to cross-dataset SOTA performance, while maintaining robustness to camera and dataset variations. These results highlight the potential of transferring IID-learned illumination priors to achieve truly sensor-agnostic color constancy.

Index Terms— Color Constancy, Intrinsic Image Decomposition

1. INTRODUCTION

Color constancy refers to the ability of the human visual system to perceive stable surface colors despite variations in scene illumination. Inspired by this perceptual capability, computational color constancy aims to estimate and compensate for the effects of illumination in digital images, thereby recovering a consistent and device-independent representation of colors.

While statistical methods for color constancy rely on low-level image statistics, deep learning-based approaches have recently achieved state-of-the-art performance due to their ability to exploit contextual information. However, this class of methods has to face some challenges.

One of the most important is camera dependency: each sensor, given its unique sensitivity across the light spectrum, renders the RAW image in its own specific color space. This implies that directly testing a CC model on an unseen camera may deteriorate its performance. Methods for effectively solving cross-camera CC have been proposed in the literature [1, 2, 3]; however, they still typically rely on training on RAW datasets, or they require additional inputs to perform CC.

Table 1: Overview of the main CC datasets.

Dataset	# images	# cameras
Gehler-Shi [4]	568	2
NUS-8 [5]	1736	8
Cube++ [6]	4890	2
INTEL-TAU [7]	7022	3

Table 2: Examples of publicly available IID datasets.

Dataset	# images
Hypersim [8]	70838
Videogames [9]	223197
Eden [10]	368663

Another potential issue related with CC is data availability, which may not be enough for training large CC methods. Table 1 summarizes the main CC datasets.

On the other hand, IID is a task in which an input image I must be decomposed as follows:

$$I = A \cdot S \quad (1)$$

where I is the input image, A is the albedo, representing the proportion of reflected light (i.e., reflectance), and S is the shading, which contains illumination intensity and, depending on the type of information that has to be considered, illumination chromaticity. Equation 1 refers to the Lambertian image formation model, which is widely used in the literature.

In contrast to CC, the field of IID has recently benefited from large-scale, camera-agnostic datasets (see Table 1 and Table 2 for a comparison). Since IID models are trained on these datasets, they can implicitly learn camera-independent priors about scene lighting.

These observations motivate a new paradigm for CC. Instead of directly learning from CC datasets, illumination priors can be transferred from the IID domain to the CC domain. In this work, we show that the recent advancements in the domain of IID datasets allow for the development of a zero-shot CC framework that does not require any task-specific training.

2. RELATED WORK

2.1. Single illuminant color constancy

Many CC methods have been proposed in the literature. Statistical learning-free methods for color constancy have been developed. Learning-based methods have been proposed as well. Convolutional networks have been successfully applied to solve the color constancy problem [11, 12, 13], estimating the illuminant chromaticity by taking into account the context of a scene and making use of its semantics. Recently, diffusion models have been exploited for performing CC [14].

Cross-camera CC methods have been previously proposed [1, 3, 2]. Afifi and Brown [1] transform the input images into a learned working space and then estimate the scene illuminant. C_5 [3] achieves cross-camera CC by taking at test time a set of raw unlabeled images taken with the same camera. CCMNet [2] pairs instead the input image with the color correction matrices of the input source sensor to reduce camera biases.

2.2. Intrinsic image decomposition

Many IID methods based on the Lambertian image formation model (See Equation 1) have been proposed in the literature. Methods based on hand-crafted features have been proposed, as well as optimization-based methods. As well as for the case of color constancy, the advent of deep-learning made it possible to learn to estimate shading and albedo by analyzing the scene semantics. Convolutional approaches have been proposed in the literature, trained in both supervised and unsupervised approaches [15, 16]. More recently, diffusion model-based methods have also been proposed.

Deep learning-based approaches have allowed not only to exploit the scene semantics to improve the decomposition performance on the Lambertian model, but also to extend the image formation model to take into account the non-Lambertian illumination effects, also called residual. In particular, the residual model is defined as follows:

$$I = A \cdot S + R \quad (2)$$

where, in this case, A is the albedo, S is the diffuse (or Lambertian) shading, and R is the residual, which contains all the non-diffuse illumination effects in the scene. Latest advancements in the field of IID have enabled us to exploit them in the domain of real-world color constancy. Careaga et al. [17] proposed a convolutional method for residual IID by creating a pipeline composed of different networks, each one predicting an intrinsic component or one of its parts.

3. PROPOSED METHOD

Our pipeline allows us to estimate global illuminant from a raw input image by leveraging albedo estimation and clustering. An overview of the proposed method is depicted in

Figure 1. The next subsections aim to explain in detail each of the steps of the pipeline.

3.1. Input normalization

RAW RGB images captured by different sensors are represented in intrinsically different color spaces, due to the specific spectral sensitivity curves of the R, G, and B channels of each camera. By contrast, the datasets used to train IID models are typically expressed in a standardized, camera-agnostic color space. This domain gap can introduce chromatic distortions in the estimated albedo, compromising the subsequent derivation of illumination.

To reduce this sensory domain gap, we apply preliminary chromatic normalization to the input image. In particular, we use a global balancing statistical method that has the sole purpose of attenuating differences due to the sensor’s color space, without introducing any specific information about the scene content. The illuminant estimated at this stage is used exclusively to normalize the image before it enters the IID model.

In our framework, we adopt Grayness Index [18], chosen for its unsupervised nature and its ability to provide robust global color correction. This step makes the input more compatible with the IID data domain, improving the estimated albedo quality and increasing the stability of the entire pipeline, while maintaining an overall learning-free approach. We mask the ColorChecker before estimating the illuminant.

3.2. Intensity masking

In this step, a binary mask is computed to exclude partially or totally clipped pixels, as well as dark pixels, during the subsequent steps. The goal is to exclude any image region whose illumination estimation might be influenced by sensor noise or sensor saturation, potentially introducing nonexistent illumination chromaticities into the prediction.

3.3. Albedo and shading estimation

This step consists of decomposing the input image using a pre-trained IID model. Any in-the-wild albedo estimation model may be employed within the proposed framework. In this case, the employed model is the albedo estimation module of the IID model from Careaga et al. [17]. Further details on training data and the used weights are provided in the supplementary material. Once the albedo is estimated, shading is derived using the original input image, following the Lambertian model in Equation 1. The result is then normalized to obtain a first version of the illumination chromaticity map. We chose not to take into account the residual model in Equation 2 as there is scarce availability of diffuse shading in IID datasets, compared to the albedo ground truth.

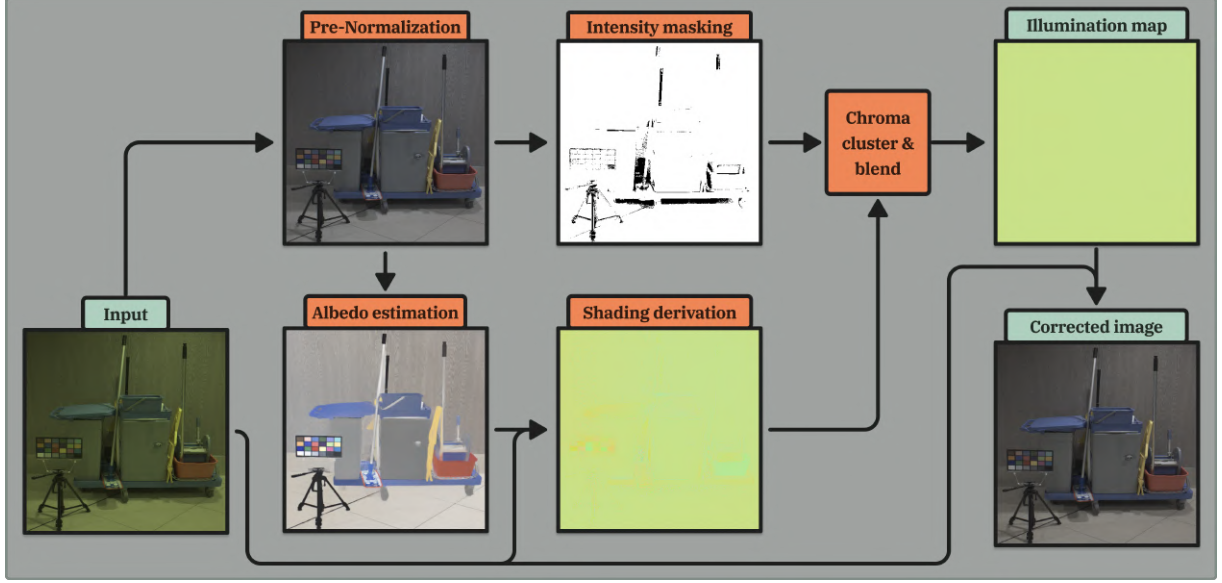


Fig. 1: Overview of the proposed method. The image is first globally balanced to mitigate differences between camera sensor color spaces, then the output is used to estimate the scene albedo. Following Equation 1, dividing the original input image by the estimated albedo will imply the image shading. Part of the shading is then selected based on the pixel intensity values of the image to exclude dark or saturated regions, and a refined shading chromaticity estimation is then obtained by blending the computed chroma clusters by computing a weighted average on each pixel. The image is then globally balanced by discounting the mean of the illumination map. Gamma correction is applied to all the images in the diagram for visualization.

3.4. Shading chroma clustering and blending

The previously obtained illumination map might contain unwanted changes in illumination due to possible albedo estimation errors, potentially leading to the introduction of small regions containing incorrect information about illumination, as well as slight but abrupt differences in illumination corresponding to changes in scene reflectance. In addition, partially clipped pixels might introduce some distortion in the illuminant chromaticity. To solve these problems, a refinement of the illumination map is achieved by clustering its chromaticities and using the clusters to generate a smoother and more precise illumination map. Clustering is not intended to explicitly segment multiple illuminants, but to suppress spurious chromaticities arising from local albedo estimation errors, thereby stabilizing the illumination estimation rather than enforcing a multi-illuminant model.

To achieve this, two steps are performed:

1. A clustering of the (R/G, B/G) chromaticities of the first output map L is performed using mean shift [19]. This clustering algorithm has been chosen because it does not require to fix the number of clusters. To make this step computationally feasible, the input to the clustering step is resized to 64×64 ;
2. A new illumination map L' is generated by blending the found illumination clusters. At the pixel i , it is

computed as a weighted combination of the found centroids:

$$L'_i = \sum_k W(k, L_i) \cdot C_k \quad (3)$$

where k is the cluster identifier, C_k is the centroid of the cluster k representing one of the candidate illumination chromaticities of the scene, and W is a weight function defined as follows:

$$W(k, L_i) = \frac{\left(\frac{\dim(k)}{\max_j \dim(j)} \right)^2}{\left(\frac{\text{dist}(L_i, C_k)}{\max_j \text{dist}(L_i, C_j)} + \epsilon \right)^2} \quad (4)$$

where $\dim(k)$ refers to the number of pixels in the k^{th} cluster, and $\text{dist}(L_i, C_k)$ represents the euclidean distance between the i^{th} pixel and the k^{th} cluster.

This weighting function assigns the contribution of each cluster to the i^{th} illumination chromaticity pixel based on how different the pixel chroma is from the centroid and how large the cluster is. In this way, small and sparse clusters tend to be considered noise rather than actual illumination, shifting those pixels towards other centroids.

Finally, the output global illuminant is computed by averaging the blended illumination map.

4. EXPERIMENTS

We assess performance on different CC datasets, as well as the robustness of our method to changes in both cameras and datasets. In addition, an ablation study is conducted to verify the effectiveness of each module in the proposed framework.

4.1. Experimental setup

The mean shift bandwidth is computed in such a way that it covers 5% of the pixels. The normalization step of our pipeline is performed on images with masked Color Checkers. Evaluation is conducted on both Gehler-Shi [4] and NUS-8 [5] datasets. We have chosen these two benchmarks as they are the most popular and allow for cross-camera evaluation performance. We compare the average RGB of the refined illumination map with the global RGB ground truths.

4.2. Evaluation metric

In order to evaluate the performance of the proposed method, the recovery angular error between the RGB estimated illuminant $\mathbf{I}^{est}$ and the RGB ground truth $\mathbf{I}^{gt}$ is computed [20]:

$$e_{\text{recovery}}(\mathbf{I}^{gt}, \mathbf{I}^{est}) = \cos^{-1} \left(\frac{\mathbf{I}^{gt} \cdot \mathbf{I}^{est}}{\|\mathbf{I}^{gt}\| \|\mathbf{I}^{est}\|} \right) \quad (5)$$

4.3. Results

Table 3 reports the average angular error for each camera subset of the NUS-8 dataset, along with the geometric mean of the average errors for each camera subset. Performance tends to be aligned across different cameras, suggesting that the method can be applied to different cameras without the need to model the sensor behaviors or fine-tune the IID backbone.

Table 4 reports cross-dataset performance comparisons on the Gehler-Shi [4] and NUS-8 [5] datasets, where the other learning-based methods have been trained on one dataset and tested on the other. Results show that our method outperforms some of the cross-camera methods ([3, 1]), except for CCM-Net [2], which achieves better results on the NUS-8 dataset [5] in this testing protocol. The main reason behind this gap in performance between CCMNet and the other methods might be the use of additional camera-specific information along with the input.

Our results on the two datasets are relatively close to each other, suggesting a certain level of generalization by the proposed method. Among all the methods, ours shows the lowest variation in angular error across datasets, aligned with GCC [14]. This might be a sign of a deeper semantic understanding of the scenes due to the exposure of the IID model to a much higher amount and variety of camera-agnostic training scenes.

Table 3: Camera-level mean angular error, expressed in degrees, of the proposed method on the NUS-8 dataset [5]. G.M. stands for geometric mean.

Camera								G.M.
1	2	3	4	5	6	7	8	
2.46	2.30	2.47	2.44	2.49	2.21	2.48	2.37	2.40

Visual results. Figure 2 reports some visual examples of input images along with the estimated albedo, the implied shading chroma, the refined shading chroma and the balanced images, along with the related ground truth maps, from the used datasets [4, 5].

4.4. Ablation studies

Ablation studies are conducted to account for the effectiveness of each of the pipeline modules. In particular, the effect of the pre-normalization and clustering are tested on the Gehler-Shi [4] and NUS-8 [5] datasets. Results are reported in Table 4.

Effect of pre-balance As already mentioned, IID datasets do not model sensor color spaces, providing images that are rendered in a standard color space; therefore, directly passing a raw image to an IID model might lead to unwanted color casts in the estimated albedo, due to the domain difference between CC and IID. Introducing this step helps to obtain a globally more precise albedo in both datasets.

Effect of clustering. Introducing the clustering step to obtain a refined illumination map helps in reducing the angular error. This is due to the fact that the estimated albedo might contain local chromaticity errors, and implying the shading can lead to abrupt chroma changes. With this step, wrong chromaticity regions can shift towards larger clusters within the scene, leading to a smoother and more correct illumination map.

5. CONCLUSIONS AND FUTURE WORK

We introduced a new way to approach CC by linking it to IID. Instead of learning illumination directly from color constancy datasets, our framework uses the priors learned by an IID model to estimate the scene illuminant, and it is able to generalize across cameras and datasets. This design allows the system to achieve illuminant estimation performance comparable to SOTA cross-camera methods in a cross-dataset setting, without any fine-tuning. Results highlight the potential of IID-based representations as a foundation for sensor-agnostic color constancy. In future work, we plan to make the approach more efficient and robust by improving the IID backbone, and to enable the pipeline for robust pixel-level multiple illuminant estimation.

Table 4: Cross-dataset evaluation on NUS-8 [5] and Gehler-Shi [4] datasets. The learning-based methods have been trained on one dataset and tested on the other. The used metric is the angular error. *Italic bold* means best in its category, while **bold** means best of all methods. (**) means that the method uses additional information along with the input image. Despite being learning-free, our method shows the smallest performance drop across datasets.

Method	Gehler-Shi [4]					NUS-8 [5]				
	Mean	Median	Tri-mean	Best 25%	Worst 25%	Mean	Median	Tri-mean	Best 25%	Worst 25%
Statistical Methods										
White-Patch	7.55	5.68	6.35	1.45	16.12	9.91	7.44	8.78	1.44	21.27
Gray-World	6.36	6.28	6.28	2.33	10.58	4.59	3.46	3.81	1.16	9.85
1st-order Gray-Edge	5.33	4.52	4.73	1.86	10.43	3.35	3.46	2.76	0.79	7.18
2nd-order Gray-Edge	5.13	4.44	4.62	2.11	9.26	3.36	2.70	2.80	0.89	7.14
Shades-of-Gray	4.93	4.01	4.23	1.14	10.20	3.67	2.94	3.03	0.99	7.75
General Gray-World	4.66	3.48	3.81	1.00	10.09	3.20	2.56	2.68	0.85	6.68
Grey Pixel	4.60	3.10	-	-	-	3.15	2.20	-	-	-
Cheng et al. [5]	3.52	2.14	2.47	0.50	8.74	2.92	2.04	2.24	0.62	6.61
LSRS [21]	3.31	2.80	2.87	1.14	6.39	3.45	2.51	2.70	0.98	7.32
Grayness Index [18]	3.07	1.87	2.16	0.43	7.62	2.91	1.97	2.13	0.56	6.67
Learning-based Methods										
Bayesian [22]	4.75	3.11	3.50	1.04	11.28	3.65	3.08	3.16	1.03	7.33
Chakrabarti [23]	3.52	2.71	2.80	0.86	7.72	3.89	3.10	3.26	1.17	7.95
FFCC [24]	3.91	3.15	3.34	1.22	7.94	3.19	2.33	2.52	0.84	7.01
SqueezeNet-FC ⁴ [12]	3.02	2.36	2.50	0.81	6.36	2.40	2.03	2.10	0.70	4.80
C ⁴ _{SqueezeNet-FC⁴} [25]	2.73	2.20	2.28	0.72	5.69	2.28	1.90	1.97	0.67	4.60
SHIE [1]	3.72	2.46	2.79	1.02	8.51	4.24	3.88	3.93	1.45	7.66
CLCC [26]	3.05	2.44	2.51	0.89	6.30	3.42	2.95	3.06	0.94	6.70
C ₅ [3] (**)	3.34	2.57	2.68	0.78	7.39	2.65	1.98	2.14	0.66	5.72
CCMNet [2] (**)	2.50	1.64	1.81	0.46	6.04	2.02	1.45	1.57	0.46	4.56
GCC [14]	2.35	2.02	2.06	0.78	4.57	2.38	2.01	2.10	0.80	4.58
Zero-Shot										
(Ours) IID-only	4.55	3.45	3.66	1.07	10.03	5.01	3.83	4.20	1.36	10.54
(Ours) IID + pre	2.51	1.88	2.04	0.61	5.45	2.70	2.15	2.31	0.76	5.54
(Ours) IID + pre + cluster	2.43	1.88	2.02	0.55	5.25	2.40	1.86	1.99	0.63	5.09

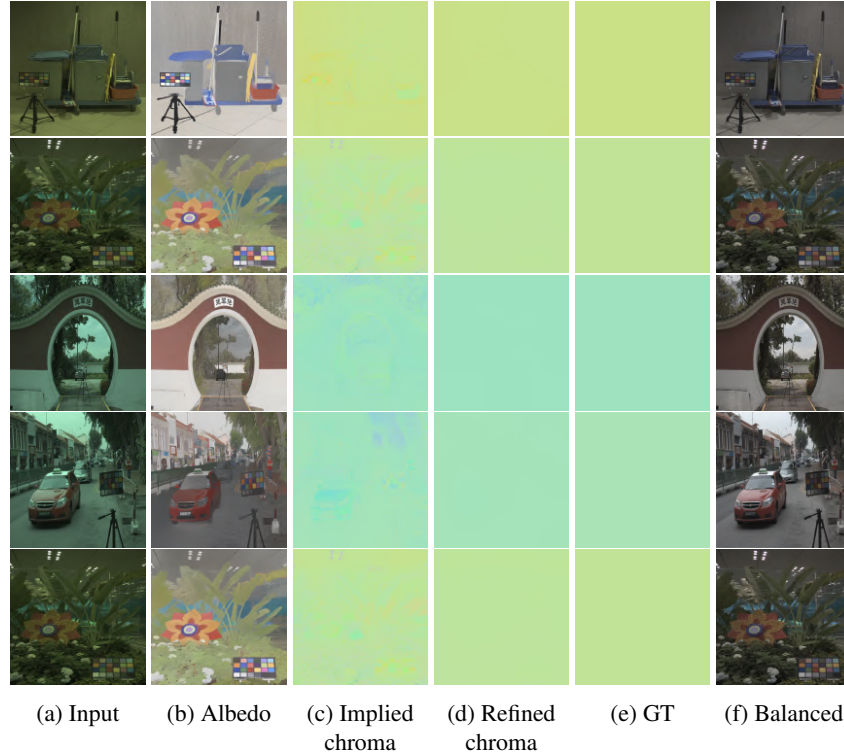


Fig. 2: Qualitative results and intermediate outputs of the proposed method on images from the Gehler-Shi [4] and NUS-8 [5] datasets. Albedo estimation removes surface appearance variations caused by illumination effects, while chromaticity clustering refines the implied shading, producing smoother illumination maps and more accurate color-balanced images. Gamma correction is applied for visualization.

6. REFERENCES

- [1] Mahmoud Afifi and M. S. Brown, “Sensor-independent illumination estimation for dnn models,” in *British Machine Vision Conference*, 2019.
- [2] Dongyoung Kim, Mahmoud Afifi, Dongyun Kim, Michael S. Brown, and Seon Joo Kim, “Ccmnet: Leveraging calibrated color correction matrices for cross-camera color constancy,” in *ICCV*, 2025, pp. 6198–6208.
- [3] Mahmoud Afifi, Jonathan T Barron, Chloe LeGendre, Yun-Ta Tsai, and Francois Bleibel, “Cross-camera convolutional color constancy,” in *Proceedings of the IEEE/CVF ICCV*, 2021, pp. 1981–1990.
- [4] Lilong Shi, “Re-processed version of the gehler color constancy dataset of 568 images,” <http://www.cs.sfu.ca/~color/data/>, 2000.
- [5] Dongliang Cheng, Dilip K Prasad, and Michael S Brown, “Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution,” *Journal of the Optical Society of America A*, vol. 31, no. 5, pp. 1049–1058, 2014.
- [6] Egor Ershov, Alexey Savchik, Illya Semenov, Nikola Banić, Alexander Belokopytov, Daria Senshina, Karlo Koščević, Marko Subašić, and Sven Lončarić, “The cube++ illumination estimation dataset,” *IEEE access*, vol. 8, pp. 227511–227527, 2020.
- [7] Firas Laakom, Jenni Raitoharju, Jarno Nikkanen, Alexandros Iosifidis, and Moncef Gabbouj, “Intel-tau: A color constancy dataset,” *IEEE access*, vol. 9, pp. 39560–39567, 2021.
- [8] Mike Roberts, Jason Ramapuram, Anurag Ranjan, Atulit Kumar, Miguel Angel Bautista, Nathan Paczan, Russ Webb, and Joshua M Susskind, “Hypersim: A photorealistic synthetic dataset for holistic indoor scene understanding,” in *ICCV*, 2021, pp. 10912–10922.
- [9] Philipp Krähenbühl, “Free supervision from video games,” in *CVPR*, 2018, pp. 2955–2964.
- [10] Hoang-An Le, Thomas Mensink, Partha Das, Sezer Karaoglu, and Theo Gevers, “Eden: Multimodal synthetic dataset of enclosed garden scenes,” in *WACV*, 2021, pp. 1579–1589.
- [11] Simone Bianco, Claudio Cusano, and Raimondo Schettini, “Color constancy using cnns,” in *CVPR workshops*, 2015, pp. 81–89.
- [12] Yuanming Hu, Baoyuan Wang, and Stephen Lin, “Fc4: Fully convolutional color constancy with confidence-weighted pooling,” in *CVPR*, 2017, pp. 4085–4094.
- [13] Firas Laakom, Jenni Raitoharju, Alexandros Iosifidis, Jarno Nikkanen, and Moncef Gabbouj, “Color constancy convolutional autoencoder,” in *SSCI. IEEE*, 2019, pp. 1085–1090.
- [14] Chen-Wei Chang, Cheng-De Fan, Chia-Che Chang, Yi-Chen Lo, Yu-Chee Tseng, Jiun-Long Huang, and Yu-Lun Liu, “Gcc: Generative color constancy via diffusing a color checker,” in *CVPR*, 2025, pp. 10868–10878.
- [15] Chris Careaga and Yağız Aksoy, “Intrinsic image decomposition via ordinal shading,” *ACM Transactions on Graphics*, vol. 43, no. 1, pp. 1–24, 2023.
- [16] Yunfei Liu, Yu Li, Shaodi You, and Feng Lu, “Unsupervised learning for intrinsic image decomposition from a single image,” in *CVPR*, 2020, pp. 3248–3257.
- [17] Chris Careaga and Yağız Aksoy, “Colorful diffuse intrinsic image decomposition in the wild,” *ACM Transactions on Graphics (TOG)*, vol. 43, no. 6, pp. 1–12, 2024.
- [18] Yanlin Qian, Joni-Kristian Kamarainen, Jarno Nikkanen, and Jiri Matas, “On finding gray pixels,” in *CVPR*, 2019, pp. 8062–8070.
- [19] Dorin Comaniciu and Peter Meer, “Mean shift: A robust approach toward feature space analysis,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, pp. 603–619, 2002.
- [20] Steven D Hordley and Graham D Finlayson, “Re-evaluating colour constancy algorithms,” in *ICPR 2004. IEEE*, 2004, vol. 1, pp. 76–79.
- [21] Shaobing Gao, Wangwang Han, Kaifu Yang, Chaoyi Li, and Yongjie Li, “Efficient color constancy with local surface reflectance statistics,” in *ECCV*. Springer, 2014, pp. 158–173.
- [22] Peter Vincent Gehler, Carsten Rother, Andrew Blake, Tom Minka, and Toby Sharp, “Bayesian color constancy revisited,” in *2008 CVPR. IEEE*, 2008, pp. 1–8.
- [23] Ayan Chakrabarti, “Color constancy by learning to predict chromaticity from luminance,” *Advances in Neural Information Processing Systems*, vol. 28, 2015.
- [24] Jonathan T Barron and Yun-Ta Tsai, “Fast fourier color constancy,” in *CVPR*, 2017, pp. 886–894.
- [25] Huanglin Yu, Ke Chen, Kaiqi Wang, Yanlin Qian, Zhaoxiang Zhang, and Kui Jia, “Cascading convolutional color constancy,” in *AAAI*, 2020, vol. 34, pp. 12725–12732.
- [26] Yi-Chen Lo, Chia-Che Chang, Hsuan-Chao Chiu, Yu-Hao Huang, Chia-Ping Chen, Yu-Lin Chang, and Kevin Jou, “Clcc: Contrastive learning for color constancy,” in *CVPR*, 2021, pp. 8053–8063.